

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 1 0 月 1 日
Date of Application:

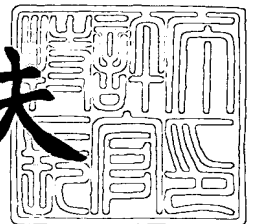
出 願 番 号 特 願 2 0 0 2 - 2 8 8 2 0 2
Application Number:
[ST. 10/C] : [J P 2 0 0 2 - 2 8 8 2 0 2]

出 願 人 株式会社日立製作所
Applicant(s):

U.S. Appln Filed 8-25-03
Inventor: M. Hara et al
mattingly Stanger & Malor
Docket NIT-394

特許庁長官
Commissioner,
Japan Patent Office

2 0 0 3 年 8 月 1 日
今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 6 1 6 4 2

【書類名】 特許願

【整理番号】 K02006611A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 原 正明

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所 ソフトウェア事業部内

【氏名】 野田 十悟

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 文書検索方法、プログラムおよびシステム

【特許請求の範囲】

【請求項 1】

計算機を用いた文書検索方法であって、
前記計算機は、ユーザから種文書の入力を受け付け、
該種文書から抽出した第 1 の特徴タームを記憶し、
文書検索処理の結果から抽出した第 2 の特徴タームを記憶し、
前記第 1 の特徴タームと前記第 2 の特徴タームとの差分を画面へ表示させることを特徴とする文書検索方法。

【請求項 2】

電子化文書の検索をするプログラムであって、
ユーザから種文書の入力を受け付け、
該種文書から抽出した第 1 の特徴タームを記憶し、
文書検索処理の結果から抽出した第 2 の特徴タームを記憶し、
前記第 1 の特徴タームと前記第 2 の特徴タームとの差分を画面へ表示させることを特徴とする電子化文書の検索をするプログラム。

【請求項 3】

電子化文書の検索システムであって、
ユーザから種文書の入力を受け付ける手段と、
該種文書から抽出した第 1 の特徴タームと文書検索処理の結果から抽出した第 2 の特徴タームとを記憶する手段と、
前記第 1 の特徴タームと前記第 2 の特徴タームとの差分を画面へ表示させる手段とを含むことを特徴とする電子化文書の検索システム。

【請求項 4】

計算機を用いた文書検索方法であって、
前記計算機は、第 1 の検索処理の結果から抽出した第 1 の特徴タームを記憶し、
第 2 の検索処理の結果から抽出した第 2 の特徴タームを記憶し、

前記第 1 の特徴タームと前記第 2 の特徴タームとを比較し、
前記比較結果を画面へ表示することを特徴とする文書検索方法。

【請求項 5】

計算機を用いた文書検索方法であって、
前記計算機は、文書検索処理の結果から抽出された特徴タームを画面へ表示し、
ユーザから前記画面に表示した特徴タームを選択する指示を受け付け、
受け付けた前記特徴タームを選択する指示を記憶することを特徴とする文書検索方法。

【請求項 6】

計算機を用いた文書検索方法であって、
前記計算機は、予め記憶装置に格納されたシソーラス分類の情報を画面へ表示し、
ユーザから画面に表示した前記シソーラス分類の情報を選択する指示を受け付け、
受け付けた前記シソーラス分類の情報を選択した指示にもとづいて文書検索処理を行うことを特徴とする文書検索方法。

【請求項 7】

計算機を用いた文書検索方法であって、
前記計算機は、ユーザから第 1 の特徴タームを受け付け、
前記第 1 の特徴タームにもとづいて検索処理を行い、該検索処理結果を画面に表示し、
ユーザから該検索処理結果にもとづいた第 2 の特徴タームの入力を受け付け、
前記第 1 の特徴タームと前記第 2 の特徴タームとを比較し、該比較結果を画面に表示することを特徴とする文書検索支援方法。

【請求項 8】

前記第 1 の特徴タームと前記第 2 の特徴タームとを比較する際に、前記第 2 の特徴タームにのみ存在する特徴タームを他の特徴タームと区別して表示することを特徴とする請求項 7 記載の文書検索支援方法。

【請求項 9】

前記第 1 の特徴タームと前記第 2 の特徴タームとを比較する際に、前記第 2 の特徴タームにのみ存在する特徴タームの重み付けを高く評価することを特徴とする請求項 7 記載の文書検索支援方法。

【請求項 10】

計算機を用いた文書検索方法であって、

前記計算機は、第 1 の特徴タームの入力をユーザから受け付け、

前記第 1 の特徴タームにもとづいて第 1 の検索処理を行い、該第 1 の検索処理の結果を画面に表示させ、

画面に表示させた前記第 1 の検索処理の結果にもとづいた第 2 の特徴タームの入力をユーザから受け付け、

前記第 1 の特徴タームと前記第 2 の特徴タームとを比較し、該比較結果にもとづいて第 2 の検索処理を行うことを特徴とする文書検索方法。

【請求項 11】

前記比較結果にもとづいて前記第 2 の検索処理を行う際に、

前記第 1 の特徴タームに含まれず前記第 2 の特徴タームにのみ含まれる特徴タームを第 3 の特徴タームとして記憶し、

前記第 3 の特徴タームの重み付けを高くし、

前記第 2 の特徴タームと前記第 3 の特徴タームとにもとづいて前記第 2 の検索処理を行うことを特徴とする請求項 10 記載の文書検索方法。

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、計算機を用いた文書検索方法、プログラムおよびシステムに関する。

【0002】**【従来の技術】**

近年、電子化された文書の増加に伴い、膨大な量の文書の中から欲しい情報をより効率よく検索するニーズが高まってきている。

従来の検索システムでは、条件（検索式）を指定し、その条件に合う文書を全て探し出してくるという手法がある。これは、ユーザの欲しい情報（文書）中で使用頻度が高いと予想される単語をもとに情報（文書）を検索すれば、その検索結果の中にユーザの欲しい情報（文書）が見つかるであろうという考えに基づいたものである。

【0003】

しかし、検索に不慣れなユーザが独力で検索効率のよい検索式を組み立てることは難しいという問題がある。

前記問題を解決する方法の一つが、概念検索と呼ばれる、検索式の代わりに文書（以下、種文書と呼ぶ）を入力して情報を検索する技術を使うことである。

【0004】

このようなユーザが入力した文書にもとづいて検索する技術としては、種文書からそれを特徴付ける単語（以下、特徴タームと呼ぶ）を抽出し、それぞれに対して適切な重みを付ける。その後、重み付けされた特徴タームに基づいて検索対象の文書に対して適合度を計算し、最後に、適合度が一定値よりも大きい文書を拾い出し、それを検索結果として表示するものがある（例えば、特許文献1）。

【0005】

また、概念検索の結果として抽出された文字列のうち、ユーザから検索結果の文書についてその要否判定を受け付け、該要否判定結果にもとづいて検索処理部（以下、概念検索トレーナ）が文字に含まれる特徴タームの重み付けを変更し、再検索するものもある（例えば、特許文献2）。

【0006】

【特許文献1】

特開2000-339346号公報（第6-7頁、第9図）

【特許文献2】

特開2001-117937号公報（第7-8頁、第15図）

【0007】

【発明が解決しようとする課題】

従来の技術には、各々以下のような問題点があった。

概念検索の問題点は、ユーザが必要としない文書が数多くヒットすることがあり、ユーザはその検索結果の一つ一つに目を通して真に欲しい文書を探すことに苦勞することがある点である。その原因の一つは、ユーザが入力する種文書にある。種文書に含まれる単語と、検索対象の文書中に含まれる単語が大きく異なる場合、概念検索は有効な特徴タームを抽出することができない。

【0008】

概念検索トレーナの問題点は、概念検索トレーナは、ユーザが要否判定した文書中に含まれる特徴タームの重み付けを自動的に変更するが、その変更が必ずしも検索精度の向上につながらない場合があることである。その原因は、ユーザが文書の要否判定するときに参考とした特徴タームと、概念検索トレーナが統計的手法により重み付けを変更する必要があると判断した特徴タームとのずれにある。

【0009】

本発明の目的は、検索に用いる特徴ターム抽出しやすくし、特徴タームをチューニングすることで検索精度の向上を図ることである。

【0010】

【課題を解決するための手段】

本発明の計算機を用いた文書検索方法であって、前記計算機は、ユーザから種文書の入力を受け付け、該種文書から抽出した第1の特徴タームを記憶し、文書検索処理の結果から抽出した第2の特徴タームを記憶し、前記第1の特徴タームと前記第2の特徴タームとの差分を画面へ表示させることを特徴とする。

【0011】

また、本発明の概念検索の検索精度に関する課題を解決する手段として以下の手段を提供する。

(1) 検索対象の文書に含まれる特徴タームを表示する手段

(2) (1) で表示された特徴タームを組み合わせる種文書として概念検索に入力する手段

また、本発明の概念検索トレーナの検索精度に関する課題を解決する手段として以下の手段を提供する。

(3) ユーザが要否判定をした検索結果の文書に含まれる特徴タームのうち重み付けを変更することが望ましいものを表示する手段

(4) (3) で表示された特徴タームに対して、ユーザが重み付けを変更する必要性の有無を入力する手段

(5) (4) でユーザが重み付けを変更する必要があると判断した特徴タームに対してだけ、重み付けを自動的に変更する手段

【0012】

【発明の実施の形態】

以下、本発明の実施例について説明する。

最初に、実施例のシステムの構成について説明する。

実施例の検索システムは、図1のようなシステム構成となっている。検索システム100は、通信回線120を介して検索するためにユーザが用いるクライアント110からアクセスされる。尚、通信回線120でなく、無線などでもよいし、その他のものを介してもよい。

検索システム100は、シソーラスジェネレータ131、概念検索エンジン（概念検索トレーナ）132、特徴タームの差分を取得する機能を含む差分取得部133、画面表示・遷移制御部134の各プログラムと、概念検索DB140、文書DB141、シソーラスDB142の各DBを含む。

尚、131～134は、それぞれ独立したプログラムでもよいし、あるプログラムの中のモジュールの機能として提供されてもよいし、その他のものでもよい。また、140～142の各データベースは、ネットワークを介して読み出し可能な記憶装置であってもよいし、その他のものでもよい。尚、特徴タームとは検索のために用いる単語を含む情報である。

【0013】

クライアント110と検索システム100は、計算機であり、発明を実施するために必要なハードウェアリソース（CPU、メモリ、記憶装置など）とソフトウェアリソース（OS、アプリケーションプログラムなど）を含んでいる。尚、クライアント110は、ブラウザや各種プログラム等を利用して、ユーザが必要とする画面の表示や各種データの入力ができるものであれば、携帯端末などでもよい。

【0014】

シソーラスジェネレータ131は、シソーラスDB142にアクセスし、シソーラス分類ごとに単語を取得する機能を持つ。概念検索エンジン132は、特開2000-339346の手法で、種文書からの特徴ターム取得、並びに検索処理を行なう機能を持つ。

【0015】

差分取得部133は、システム起動から特徴ターム差分取得部133の呼び出しまでに行われた任意の2回の検索処理で使われた特徴タームの差分を取得する機能を持つ。尚、ある検索で用いる特徴タームと別な検索で用いる特徴タームとをそれぞれ記録装置に格納しておき、差分取得部133がそれらの特徴タームの差分を取得する機能を含んでもよい。画面表示・遷移制御部134は、検索で使用する画面とその遷移を制御する機能を持つ。

【0016】

また、概念検索DB140には概念検索処理に使用するインデックスが、文書DB141には検索対象の文書が、シソーラスDB142にはシソーラス分類された単語群が、それぞれ格納されている。

尚、シソーラスとは、情報を検索する場合において、検索するためのキーワードの示す範囲、検索するためのキーワードと関連語の類似・対立・包含関係などを記述したものを意味する。

【0017】

各DB140～142に関しては、プログラムと同一のサーバ内ではなく、ネットワーク上に別サーバを設けてそこに格納しても良い。

実施例の検索システムの操作手順に従って、処理の流れを図2に従って説明する。

【0018】

実施例では、図2の流れに従って検索を行う。ステップ210において、シソーラスジェネレータ131は、シソーラスDB142に格納されたシソーラスのデータを読み出す。ステップ220は、ユーザから、検索するための単語の入力を受け付ける処理である。ステップ221で、ユーザは単語選択画面（図3）を用いて、これ

から検索する文書の内容に近いものをシソーラス分類の中から選択する。

【 0 0 1 9 】

ステップ222で、ステップ211の処理で選択した単語にもとづいて、ユーザは種文書編集画面（図4）を用いて種文書を作成する。ユーザの種文書作成後、ステップ230において、概念検索エンジン132は、概念検索処理を行う。ステップ240で、ステップ230の結果を概念検索トレーナ画面（図5）に出力する。

【 0 0 2 0 】

ステップ250で、ステップ222の処理における種文書編集画面（図4）の表示でユーザから選択または追加入力された単語（第1の特徴ターム）と、ステップ240の処理における概念検索トレーナ画面（図5）でユーザから選択された文書から抽出された単語（第2の特徴ターム）とを比較し、特徴タームの差分取得の処理を行う。

【 0 0 2 1 】

ステップ260で、ユーザによる検索結果の取舍選択後、ステップ230の概念検索処理の段階では存在しなかった特徴タームを明確にし、ステップ270の概念検索処理で使用しようとしている特徴タームを選択画面（図6）に表示する。つまり、ステップ260の処理で前述のステップ250の処理で抽出した特徴タームを表示する。また、ステップ260で、ユーザは、検索したい内容とあまり関係のない単語を、次のステップ270の概念検索処理で使用しない特徴タームとして排除することができる。また、ステップ260の処理でユーザから選定された特徴タームを格納し、次の検索のために用いる特徴ターム（ステップ240で画面に出力される特徴ターム）としてもよい。特徴タームの選定後、ステップ270の概念検索処理を行う。

【 0 0 2 2 】

ステップ280で、ステップ270の結果をトレーニング結果画面（図7）に表示する。検索結果が得られた場合はシステムを終了し、もう一度検索を行なう場合は、ステップ240の概念検索トレーナ画面（図5）に戻り、検索結果が得られるまで繰り返す。

【 0 0 2 3 】

尚、上述した画面の表示は、クライアント110の計算機で稼動するWebブラウザ等のプログラムを用いてユーザへ見せてもよい。また、その他の方法で、クライアント110の計算機を利用して、検索システム100へアクセスして、検索処理に必要な処理を実行してもよい。

【0024】

以下、図3～図7の画面表示例と、図8～図11で示したフローチャートの例を用いて各処理の詳細を説明する。

システムを起動すると、画面表示・遷移制御部134が図3のような単語選択画面300を表示させる。

尚、ユーザへ見せる表示画面は、検索システム100をWebブラウザで表示可能な形式のファイルとして検索システム100の記憶装置に格納し、クライアント110で稼動するWebブラウザのプログラムがネットワークを介して検索システム100へアクセスし、図で示したようなページを表示してもよい。

【0025】

表示窓310には、シソーラスジェネレータ131が、シソーラスDB142から取得したシソーラス分類の情報を表示する。ユーザは、この中から検索したい情報と関係が深いと思う単語群を選択し、決定ボタン320を押す。

【0026】

ユーザからの決定ボタン320が押下の指示を受信した後、図4のような種文書編集画面400を表示する。ここでは、選択した単語群が種文書編集エリア410にあらかじめ入力されており、ユーザは種文書編集エリア410で単語を追加／削除したり、文章にしたりして種文書を作成する。ユーザは、種文書作成が終わると検索開始ボタン420を押す。検索開始ボタン420を押すと、作成された種文書で概念検索を開始する。この過程で生成される第1の特徴ターム（以下、特徴ターム（1））は、このとき検索システム100内の記憶装置に保存しておく。

【0027】

システムが起動され、ユーザが入力した種文書を受付けて、該受け付けた種文書にもとづいて概念検索をし、該受け付けた種文書を格納するまでの処理を以下に示す（図8フローチャート1）。

【 0 0 2 8 】

図 8 は、単語選択画面・種文書編集画面の画面表示の処理の流れを示すフローチャートである。

ステップ801で、シソーラスジェネレータ131が、シソーラスDB142にアクセスして、シソーラスDBに記憶されているシソーラスのデータを読み出す。

ステップ802で、画面表示・遷移制御部134が、図 3 の単語選択画面300を表示する。表示窓310中に、読み出したシソーラス分類を表示する。ユーザは、表示されたシソーラス分類の中から検索したい内容に近いものを選択する。

【 0 0 2 9 】

ステップ803で、ユーザが決定ボタン320が押下すると、画面表示・遷移制御部134が、図 4 の種文書編集画面400を表示する。種文書編集エリア410に単語群を表示する。

ステップ804で、種文書編集エリア410でユーザが種文書を編集又は作成する。

【 0 0 3 0 】

ステップ805で、ユーザが検索開始ボタン420を押下すると、該検索開始の指示を受け付けた概念検索エンジン132が、作成された種文書から特徴タームを抽出する。抽出された特徴ターム（特徴ターム（1））は、一時的な記憶域に記憶する。

ステップ806で、抽出された特徴タームを利用して、概念検索エンジンによる概念検索処理を開始する。

【 0 0 3 1 】

図 8 および図 4 を用いて説明した概念検索処理が終了した後の処理を図 9 および図 5 を用いて説明する。

概念検索処理が終了すると、図 5 のような概念検索トレーナ画面500を表示し、検索結果を概念検索トレーナ窓510に表示する。

次に、検索結果に対してトレーニングを行なう。まず、概念検索の結果として順位付けして表示した文書に対し、有用な文書とそうでない文書とをユーザが振り分ける。具体的には、ユーザが、必要な内容に近い文書には○を、あまり関係のない文書には×を概念検索トレーナ窓510の中の○×記入欄530に入力する。こ

の後、○×決定ボタン520を押すと、特徴タームの再評価の処理を開始する。

【0032】

この再評価後に生成された第2の特徴ターム（以下、特徴ターム（2））は保存し、特徴ターム（1）と比較を行なう。具体的には、差分取得プログラム133が、特徴ターム（1）には存在せず、特徴ターム（2）で新たに出現した単語を取得する。概念検索トレーナ画面500表示からここまでの処理を以下に示す（図9フローチャート2）。

【0033】

図9は、概念検索トレーナの表示画面の遷移の処理を示すフローチャートである。

ステップ901で、画面表示・遷移制御部134が概念検索トレーナ画面500を表示する。検索結果を概念検索トレーナ窓510に表示する。

ステップ902で、ユーザが検索結果の文書に対して、有用なものには○、不要なものには×を付ける。ユーザが○×決定ボタン520を押すとステップ903へ進む。

【0034】

ステップ903で、画面表示・遷移制御部134が、○の付いた文書から抽出される特徴タームの重みは高く、×の付いた文書から抽出される特徴タームの重みは低く、特徴タームの重み付けを再評価する処理を実行する。尚、特徴タームの重み付けを再評価する処理とは、ユーザからの入力指示にもとづいて、特徴タームに対応づけて格納された重み付けの情報を変更する処理等を含む。ここで再抽出された特徴タームを保存する（特徴ターム（2））。

【0035】

ステップ904で、差分取得プログラム133が、特徴ターム（1）には存在せず、特徴ターム（2）には存在する単語（特徴ターム（3））を取得する。

【0036】

特徴タームの差分取得処理が終了すると、図6のような特徴ターム選択画面600を表示する。特徴ターム選択窓610に特徴ターム（2）を表示するが、このとき特徴ターム（3）に属する単語を他の単語と区別して表示する（本実施例の図6

では文字を大きくしている）。この表示処理により、自らの○×の付加で特徴タームに加わった単語を新たな検索概念としてユーザに認知させ、必要に応じて検索対象分野の軌道修正を可能にする効果がある。

【 0 0 3 7 】

ユーザは、次の検索に必要ないと判断した単語（次回のトレーニングで特徴タームとして使用しない単語）の○×付加欄640に×をつける。（デフォルトで全ての単語に○が付いている。）トレーニング処理を行う前にこうして特徴タームの取捨選択を行なうことで検索の精度の向上をはかる。

【 0 0 3 8 】

ユーザが、画面に表示されたトレーニングボタン620を押すと、概念検索エンジン132が、○の付いた単語群を種文書として受け付け、該受け付けた単語群を種文書として概念検索処理を開始する。

ユーザが、画面に表示されたキャンセルボタン630を押すと、一つ前の概念検索トレーナ画面500に戻り、各文書への○×付加をやりなおすことができる。特徴ターム選択画面600の表示からここまでの処理を以下に示す（図 1 0 フローチャート 3）。

【 0 0 3 9 】

図 1 0 は、特徴ターム選択を行う画面表示の遷移の処理を示すフローチャートである。

【 0 0 4 0 】

ステップ1001で、画面表示・遷移制御部134が、特徴ターム選択画面600を表示する。特徴ターム選択窓610に特徴ターム（2）を表示する。ここで、特徴ターム（3）に属する単語は他の単語と区別して表示する。○×記入欄640にはすべて○を記入しておく。

ステップ1002で、ユーザは、特徴ターム選択窓610で、検索したい情報にあまり関係がないと思う単語に×をつける。

【 0 0 4 1 】

ステップ1003で、ユーザが画面に表示されたトレーニングボタン620を押すと、概念検索エンジン132が、クライアント110から○の付いた単語を種文書として

入力を受け付け、該入力として受付けた単語を種文書として概念検索処理を開始する。

ステップ1004で、キャンセルボタン630を押すと、概念検索トレーナ画面500に戻る。

【0042】

検索結果は図7のトレーニング結果画面700のトレーニング結果表示窓710に表示する。新たに順位付けられた文書の左側（順位上下表示欄740）にはその文書の順位があがったか下がったかを示す矢印を表示する。尚、文書中に含まれる特徴タームの数や文書中に含まれる特徴タームの重み付け等にもとづいて文書の順位の決定してもよいし、その他の方法で文書の順位付けを行ってもよい。

【0043】

ユーザはこの検索結果を見て、終了する場合は終了ボタン730を押す。もう一度検索する場合は再検索ボタン720を押す。再検索ボタン720を押すと、概念検索トレーナ画面500に遷移する。トレーニング結果画面700の表示からここまでの処理を示す（図11フローチャート4）。

【0044】

図11は、トレーニング結果の画面表示の遷移処理を示すフローチャートである。

ステップ1101で、画面表示・遷移制御部134が、トレーニング結果画面700を表示する。トレーニング結果表示窓710に新たに順位付けられた文書を表示し、順位上下表示欄740に各文書について前回の検索結果から順位が上がったか下がったかを示す矢印を表示する。

ステップ1102で、終了ボタン730が押された場合は検索システムを終了する。

【0045】

ステップ1103で、再検索ボタン720が押された場合は、画面表示・遷移制御部134の制御により概念検索トレーナ画面500の表示処理（ステップ901）に遷移する。

以降、ユーザに満足のいく検索結果が出るまで、文書への○×付加（ステップ901）から検索結果出力（ステップ1101）までの処理を繰り返す。

【0046】

本発明を利用することで、検索対象の文書中に含まれる特徴タームを使いながら種文書を作ることができるので、概念検索の検索精度を高めることができる。

また、明確に検索分野を絞って概念検索トレーナを使用して検索する場合、重み付けを変更する特徴タームをユーザが直接指定する本手法を付加的に使用することによって、従来よりも少ない検索回数で必要とする文書を取得することが可能になる。

【0047】

また、広範囲に情報を求める目的で検索する場合、前回の検索では抽出されず、今回新たに抽出された特徴タームをユーザに明示し、新たな概念として次の検索に取り入れさせることで、多様な情報の検索を可能にする効果がある。

従来、概念検索において、ユーザが独力で有効な種文書を作成するのは難しい。また、概念検索トレーナは、特徴タームの重み付けを自動的に変更するが、その変更が必ずしも検索精度の向上につながらない場合がある。

しかし、本発明では、最初の検索では、シソーラスを使用してユーザの種文書作成を支援し、2回目以降は、新たに抽出された特徴タームをユーザに表示し、検索精度を上げるために種文書を調整できるユーザインターフェイスを提供するため、検索精度を上げるための支援をすることができる。

【0048】

たとえば、予め記憶装置に格納されたシソーラス分類の情報を画面へ表示させ、ユーザは該表示にもとづいて特徴ターム又は種文書の指示の入力をするため、ユーザは新たに単語を入力しなくても済むので、ユーザにとって使い勝手の良い検索方法となる。また、すでに行った検索結果から特徴タームを抽出し、画面に表示することで、ユーザは該画面に表示された特徴タームに基づいて、次回に検索するときに用いる特徴タームの指示の入力や重要だと思われる単語を選択して入力し、該ユーザからの指示を記憶することで検索結果を次回の検索へ反映させることも可能となる。

このように、ユーザが種文書や特徴タームの選択又は調整（チューニング）をすることで、検索のもととなる情報をユーザの要望に応じて、より詳細に作成す

ることができ、検索の結果から重要な情報や検索の上で必要となる特徴タームを選別することで、検索精度を高めることが可能となる。

【0049】

また、本発明では、文書中の特徴タームを用いて作成した検索処理前の特徴タームと検索処理の結果から抽出した特徴タームとを比較し又は差分を抽出し、該抽出した結果を次の検索処理のために用いる特徴タームに反映させることができるので、概念検索の検索精度を高めることができる。

【0050】

また、本発明を用いて、複数の検索処理の結果から抽出した特徴タームを比較し、比較した結果を次の検索の特徴タームに反映させてもよい。

【0051】

また、本発明を用いて広範囲に情報を求める目的で検索する場合、前回の検索では抽出されず、今回新たに抽出された特徴タームをユーザに明示し、新たな概念として次の検索に取り入れさせることで、多様な情報の検索が可能になる。

【0052】**【発明の効果】**

以上説明したように本発明を用いることで、検索に用いる特徴タームをチューニングし、検索精度の向上を図ることができる。

【図面の簡単な説明】**【図1】**

本発明の実施例の構成の例を示した図である。

【図2】

実施例の画面遷移と処理内容の例を示した図である。

【図3】

単語選択画面の一例を示した図である。

【図4】

種文書編集画面の一例を示した図である。

【図5】

概念検索トレーナ画面の一例を示した図である。

【図 6】

特徴ターム選択画面の一例を示した図である。

【図 7】

トレーニング結果画面の一例を示した図である。

【図 8】

単語選択画面・種文書編集画面の流れを示すフローチャートの例である。

【図 9】

概念検索トレーナ画面の流れを示すフローチャートの例である。

【図 10】

特徴ターム選択画面の流れを示すフローチャートの例である。

【図 11】

トレーニング結果画面の流れを示すフローチャートの例である。

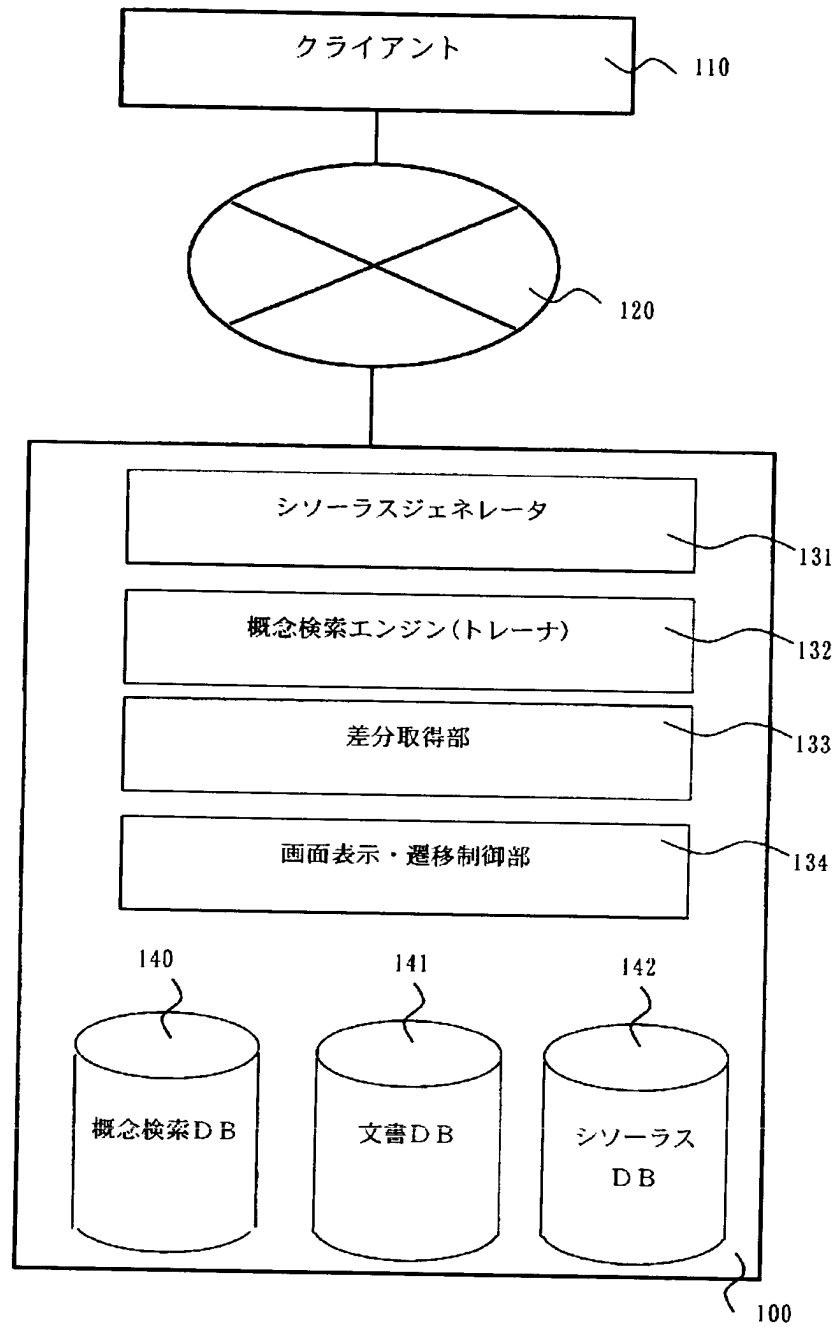
【符号の説明】

- 100 サーバ
- 110 クライアント
- 120 通信回線
- 131 シソーラスジェネレータ
- 132 検索エンジン
- 133 差分取得部
- 134 画面表示・遷移制御部
- 140 概念検索DB
- 141 文書DB
- 142 シソーラスDB

【書類名】 図面

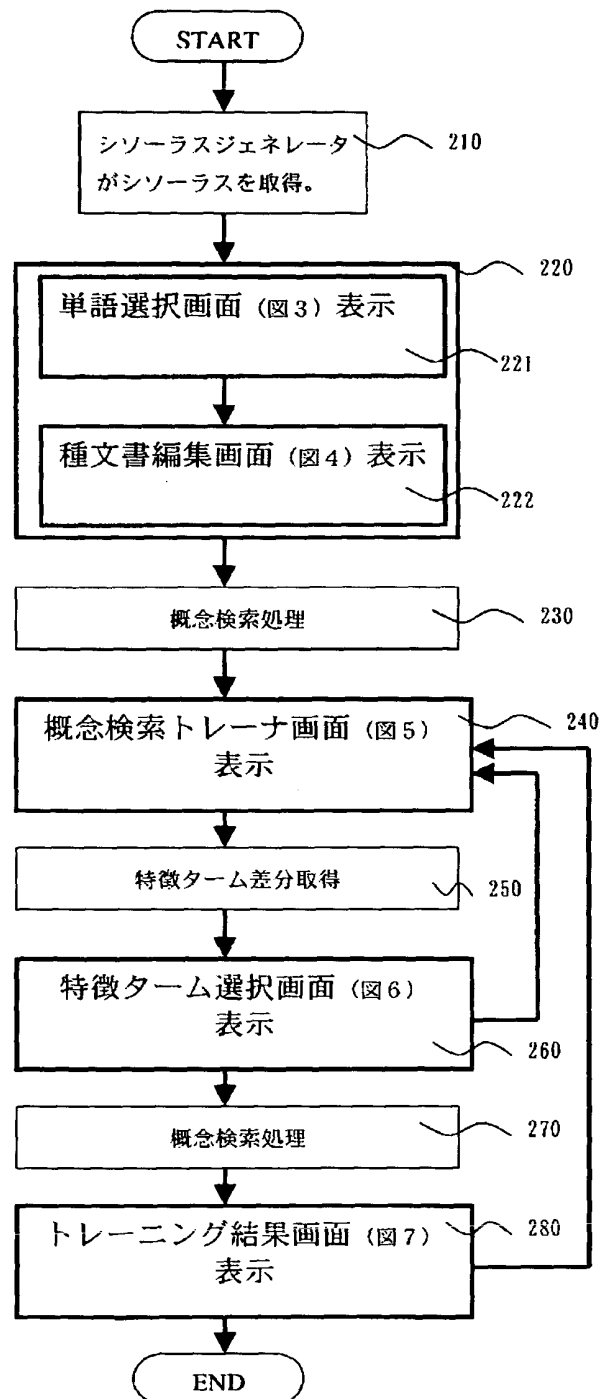
【図 1】

【図 1】



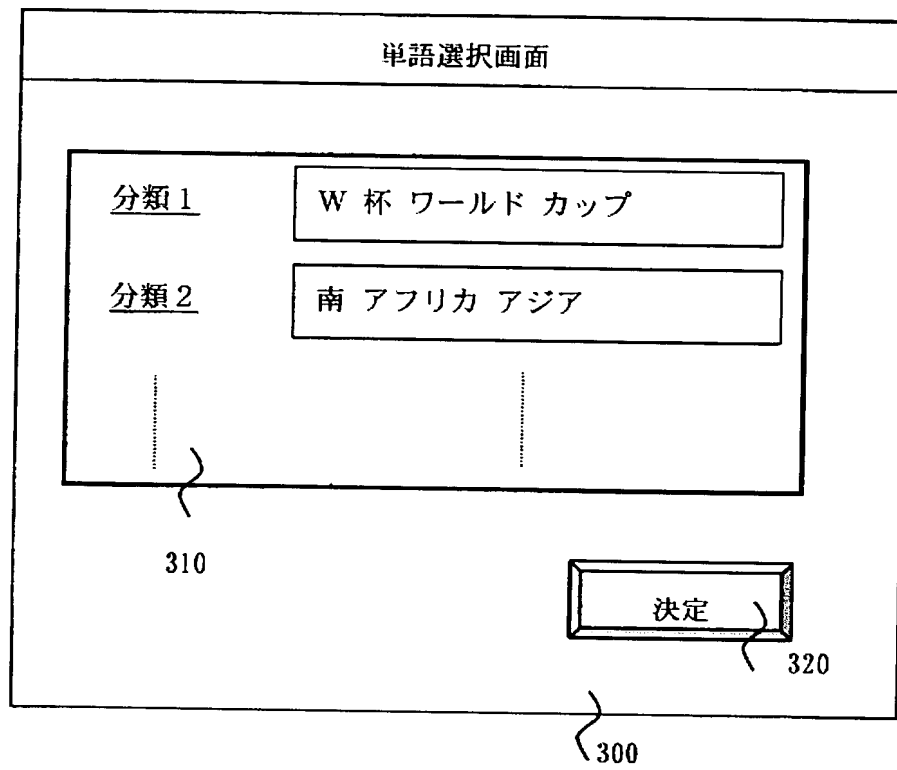
【図 2】

【図 2】



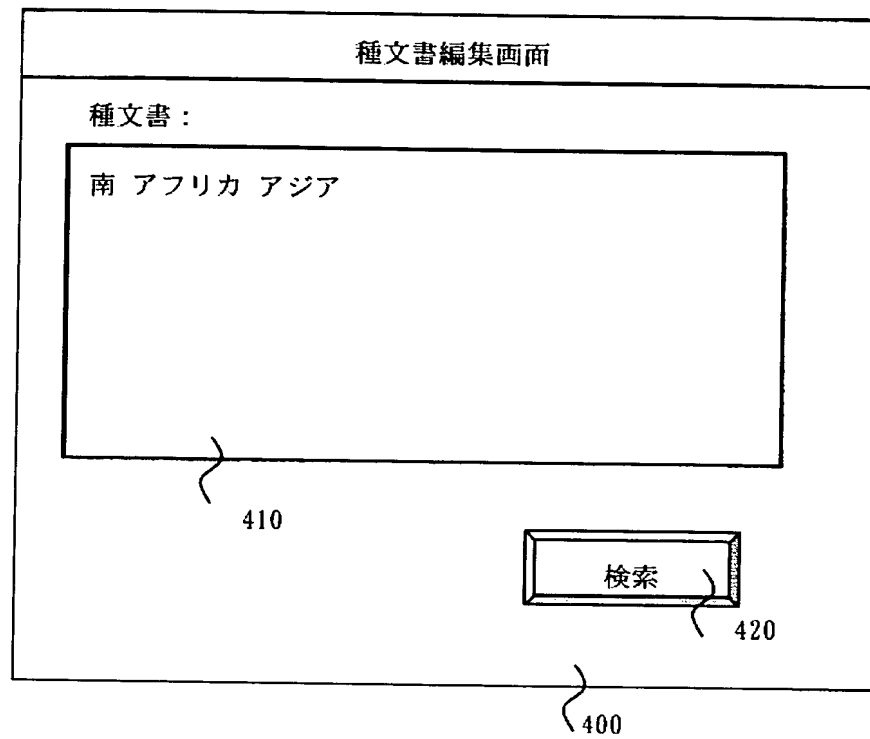
【図 3】

【 図 3 】



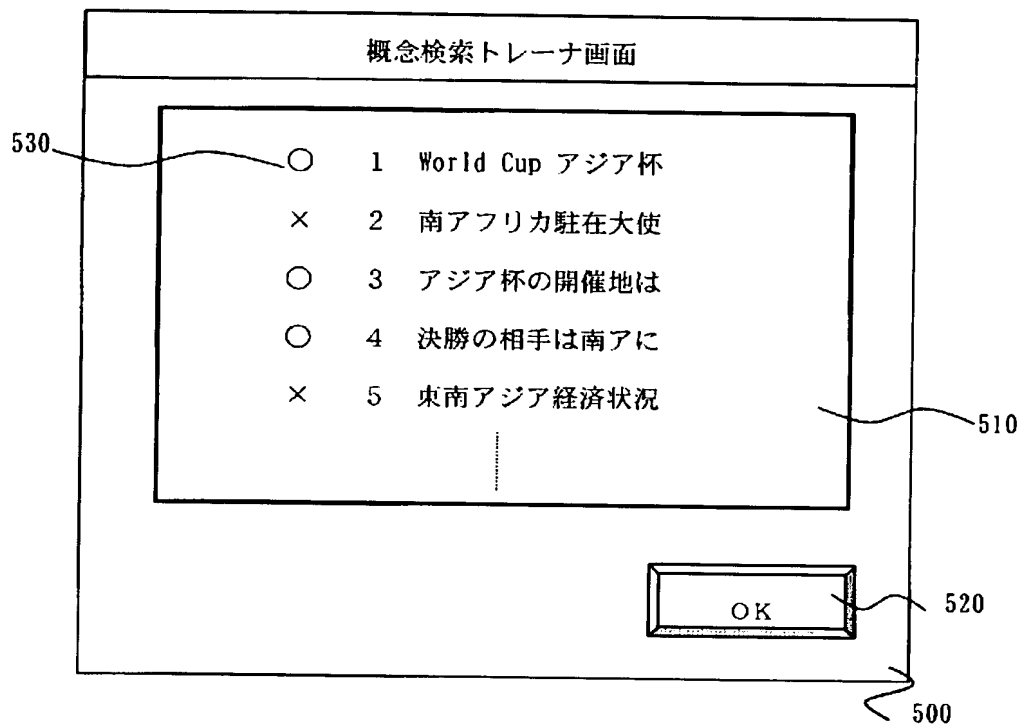
【図 4】

【 図 4 】



【図 5】

【 図 5 】



【図 6】

【図 6】

特徴ターム選択画面

<input type="radio"/>	アジア	<input type="checkbox"/>	開催地
<input type="radio"/>	決勝	<input type="checkbox"/>	相手
<input type="checkbox"/>	南		

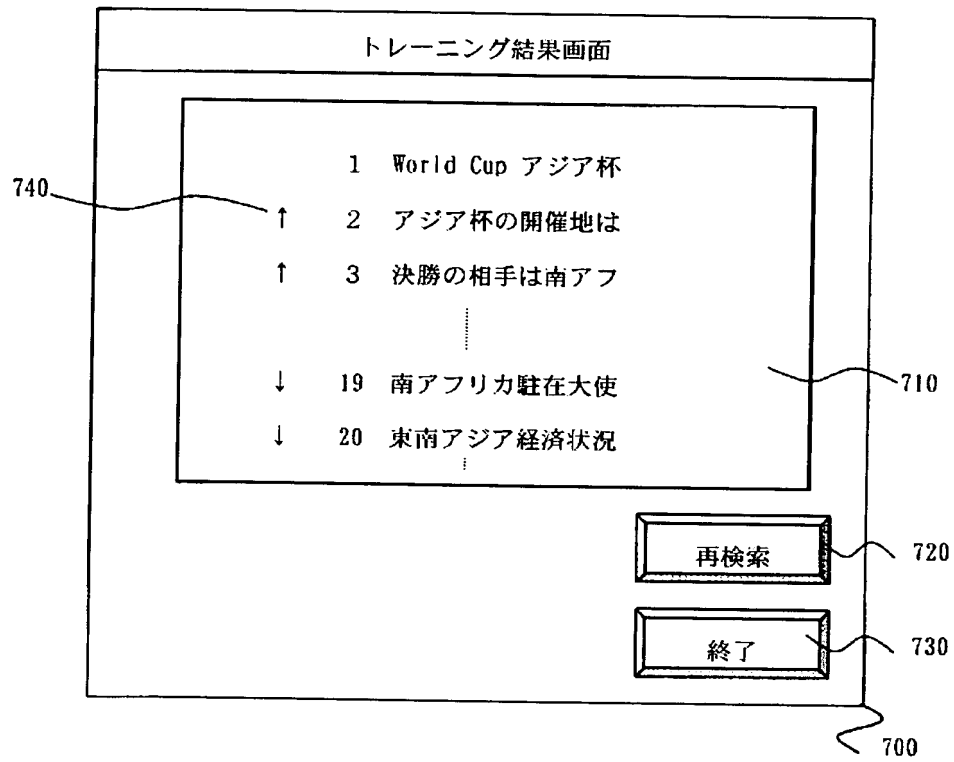
トレーニング

キャンセル

640 610 620 630 600

【図 7】

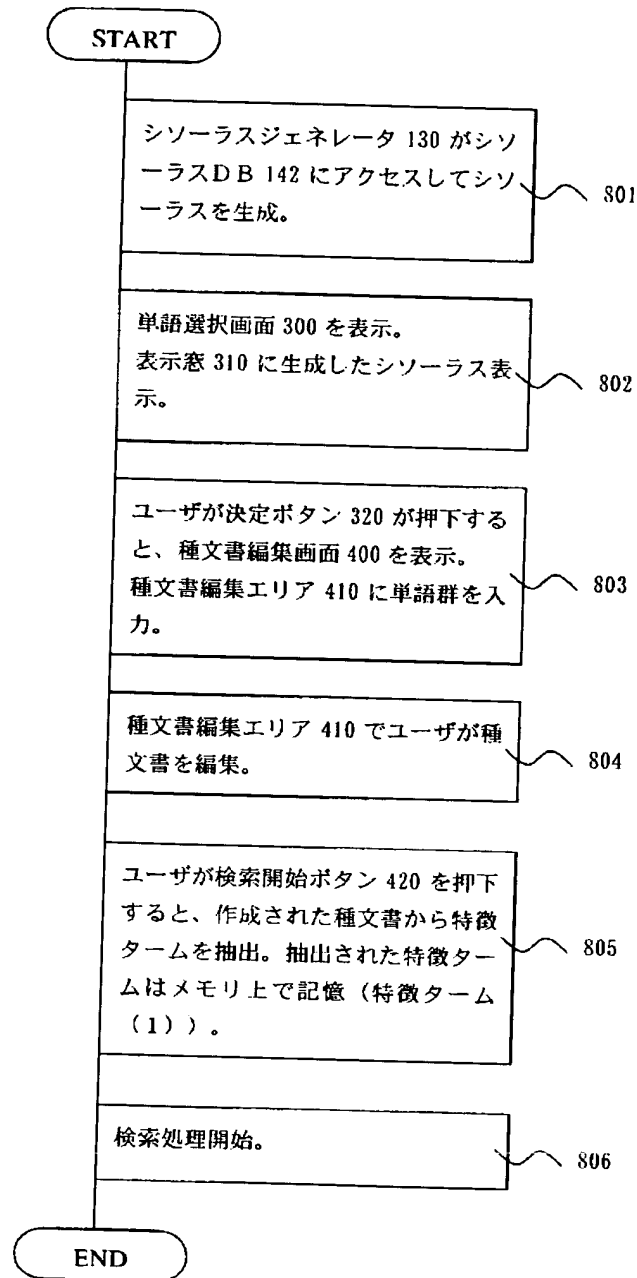
【 図 7 】



【図 8】

【図 8】

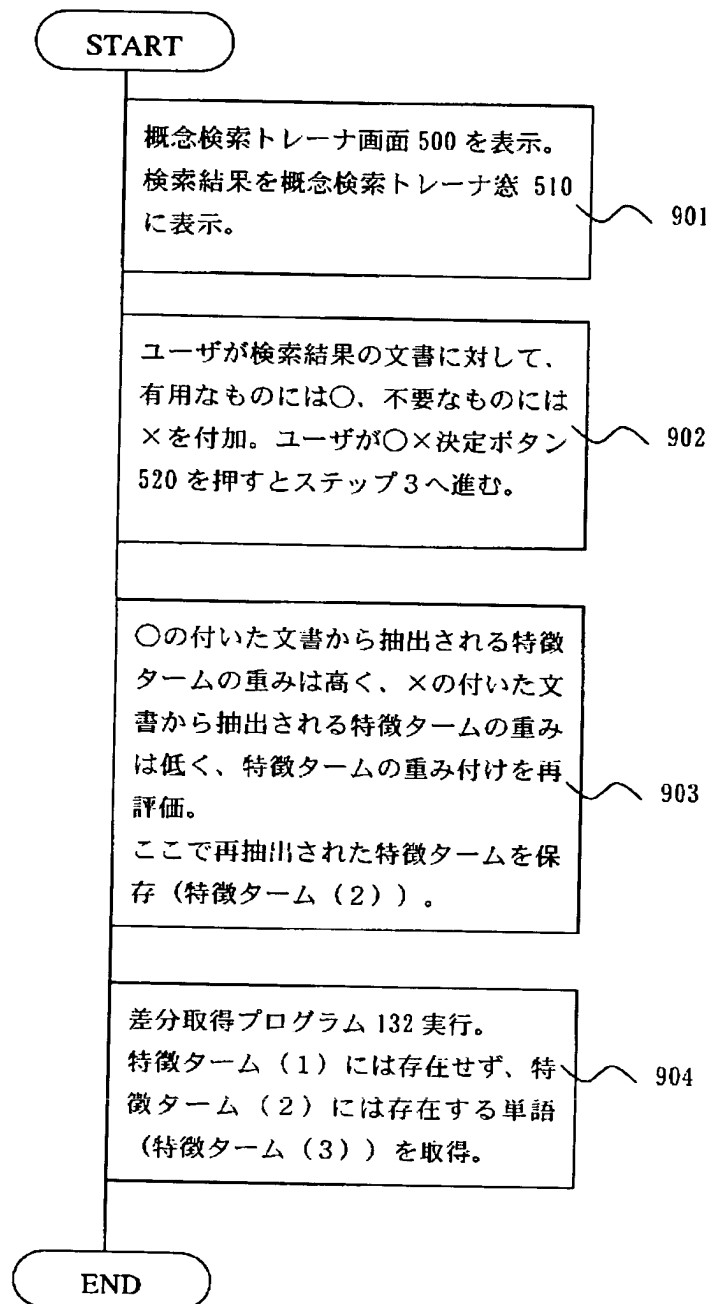
フローチャート 1 単語選択画面・種文書編集画面の流れ



【図 9】

【図 9】

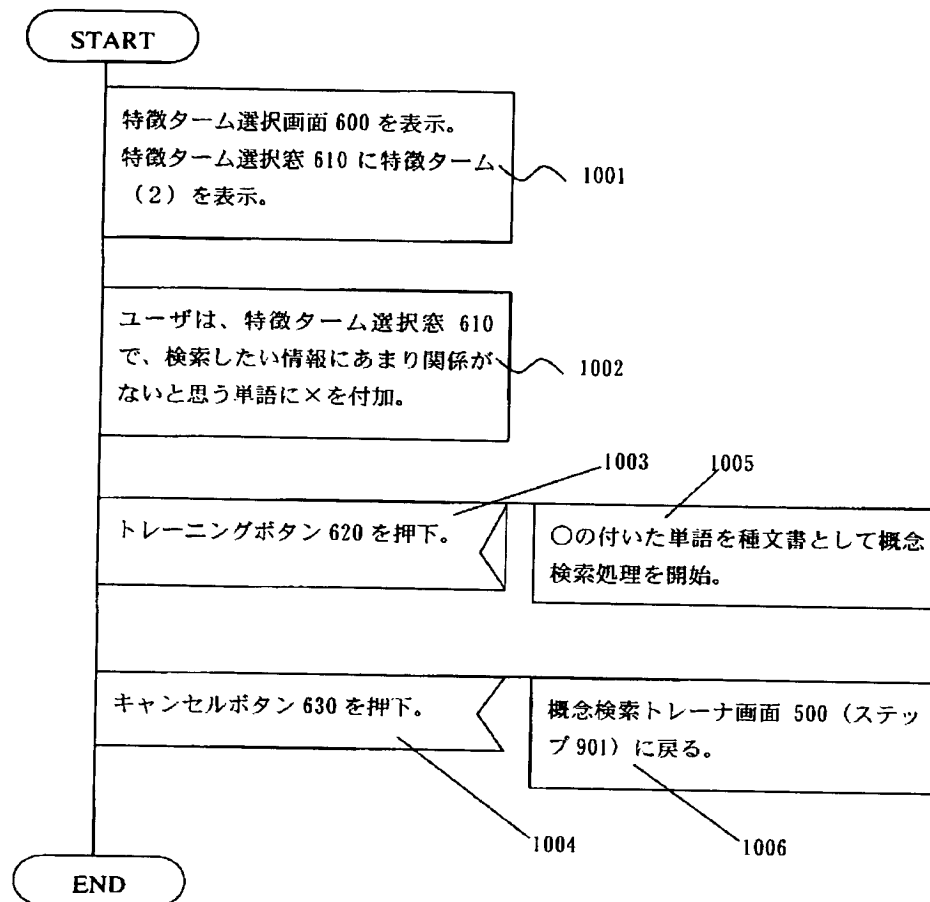
フローチャート 2 概念検索トレーナ画面の流れ



【図 10】

【図 10】

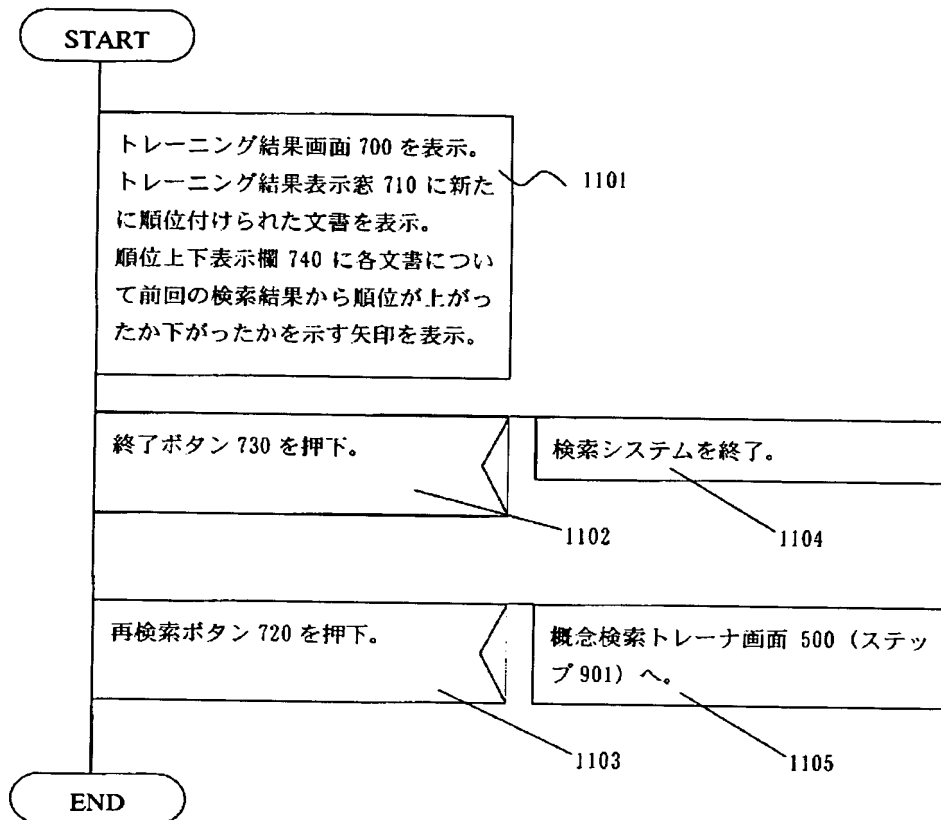
フローチャート 3 特徴ターム選択画面の流れ



【図 11】

【図 11】

フローチャート 4 トレーニング結果画面の流れ



【書類名】 要約書

【要約】

【課題】

概念検索において、ユーザが独力で有効な種文書を作成するのは難しい。また、概念検索トレーナは、特徴タームの重み付けを自動的に変更するが、その変更が必ずしも検索精度の向上につながらない場合がある。

【解決手段】

最初の検索では、シソーラスを使用してユーザの種文書作成を支援する。2回目以降は、新たに抽出された特徴タームをユーザに表示し、検索精度を上げるために種文書を調整できるユーザインターフェイスを提供する。

【選択図】 図2

認定・付加情報

特許出願の番号	特願 2 0 0 2 - 2 8 8 2 0 2
受付番号	5 0 2 0 1 4 7 2 9 1 8
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 4 年 1 0 月 2 日

＜認定情報・付加情報＞

【提出日】	平成14年10月 1日
-------	-------------

次頁無

出証特 2 0 0 3 - 3 0 6 1 6 4 2

特願 2002-288202

出願人履歴情報

識別番号

[000005108]

1. 変更年月日

1990年 8月31日

[変更理由]

新規登録

住所

東京都千代田区神田駿河台4丁目6番地

氏名

株式会社日立製作所